

UNIVERSITY OF TECHNOLOGY SYDNEY  
Faculty of Engineering and Information Technology

**Resource Management of Fog Computing in  
Future Networks**

by

**Xinchen Lyu**

A THESIS SUBMITTED  
IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE

**Doctor of Philosophy**

Sydney, Australia

2018

## Certificate of Authorship/Originality

I, Xinchun Lyu declare that this thesis, is submitted in fulfilment of the requirements for the award of doctor of philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney. This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree except as fully acknowledged within the text. This thesis is the result of a research candidature jointly delivered with Beijing University of Posts and Telecommunications as part of a Collaborative Doctoral Research Degree. This research is supported by the Australian Government Research Training Program.

Production Note:

Signature: Signature removed prior to publication.

Date: 11 April 2019

# ABSTRACT

## Resource Management of Fog Computing in Future Networks

by

Xinchen Lyu

By enabling task offloading among edge network nodes (such as base stations, switches, and routers), fog computing is able to process the computationally demanding tasks at the point of capture. It has the potential to provide low-latency services, increase network capacity, and relieve network congestions. In this thesis, we focus on three exemplary scenarios of fog computing, including (1) single-cell multiuser fog computing to jointly optimize task offloading and resource allocation for different applications with heterogeneous quality of service requirements; (2) fog computing among selfish devices to design incentive mechanism and efficient management of task offloading, processing, and result retrieving; and (3) fog computing across large-scale edge cloud for scalable and distributed resource management in the presence of a large number of geo-distributed edge servers.

We present five new approaches to address the challenges for efficient and scalable fog computing in the three scenarios. The first three approaches are for the first scenario, i.e., single-cell multiuser fog computing, for three different types of applications, including delay-tolerant, delay-sensitive, and data-partition tasks. The fourth approach is for the second scenario, where distributed tit-for-tat mechanism is proposed to incentivize the cooperation of selfish devices. The fifth approach is for the last scenario, where collaborative regions are created for preventing tasks being offloaded beyond the vicinity of the point of capture in large-scale networks.

# Acknowledgements

First, I would like to thank my principal supervisor, Prof. Ren Ping Liu, for his guidance, encouragement, enthusiasm, and technical comments through all years of my research. I also would like to express my deep gratitude to my co-supervisors, Dr. Wei Ni and Prof. Eryk Dutkiewicz, for their valuable advice and guidance of my research.

Special thanks to Beijing University of Posts and Telecommunications (BUPT), Beijing, China, and University of Technology Sydney (UTS) for providing a scholarship and other financial support for my study.

To my family, thank dad and mom for your love and encouragement to help me complete this study. I have been incredibly fortunate to have love and support from such family. Finally, to my wife, thank you for encouraging, enduring and inspiring me through the years. Love you always.

Xinchen Lyu  
Sydney, Australia, 2018.

# List of Publications

## Journal Papers

- J-1 . X. Lyu, C. Ren, W. Ni, H. Tian, R. P. Liu, Y. Jay Guo. “Multi-timescale Decentralized Online Orchestration of Software-Defined Networks” in **IEEE Journal on Selected Areas in Communications**, 2018, DOI: 10.1109/JSAC.2018.2871310.
- J-2 . X. Lyu, C. Ren, W. Ni, H. Tian, R. P. Liu. “Distributed Optimization of Collaborative Regions in Large-Scale Inhomogeneous Fog Computing” in **IEEE Journal on Selected Areas in Communications**, vol. 36, no. 3, pp. 574-586, March 2018.
- J-3 . X. Lyu, W. Ni, H. Tian, R. P. Liu, X. Wang, G. B. Giannakis, A. Paulraj. “Optimal Schedule of Mobile Edge Computing for Internet of Things Using Partial Information” in **IEEE Journal on Selected Areas in Communications**, vol. 35, no. 11, pp. 2606-2615, Nov. 2017.
- J-4 . X. Lyu, H. Tian, L. Jiang, A. Vinel, S. Maharjan, S. Gjessing, Y. Zhang. “Selective Offloading in Mobile Edge Computing for the Green Internet of Things” in **IEEE Network**, vol. 32, no. 1, pp. 54-60, Jan.-Feb. 2018.
- J-5 . X. Lyu, W. Ni, H. Tian, R. P. Liu, X. Wang, G. B. Giannakis, A. Paulraj. “Distributed Online Optimization of Fog Computing for Selfish Devices with Out-of-Date Information” in **IEEE Transactions on Wireless Communications**, 2018, DOI: 10.1109/TWC.2018.2869764.
- J-6 . X. Lyu, H. Tian, W. Ni, Y. Zhang, P. Zhang, R. P. Liu. “Energy-Efficient Admission of Delay-Sensitive Tasks for Mobile Edge Computing” in **IEEE Transactions on Communications**, vol. 66, no. 6, pp. 2603-2616,

June 2018.

- J-7 . X. Lyu, H. Tian, C. Sengul, P. Zhang. “Multiuser Joint Task Offloading and Resource Optimization in Proximate Clouds” in ***IEEE Transactions on Vehicular Technology***, vol. 66, no. 4, pp. 3435-3447, April 2017.
- J-8 . X. Lyu, H. Tian, W. Ni, R. P. Liu, P. Zhang. “Adaptive Centralized Clustering Framework for Software-Defined Ultra-Dense Wireless Networks” in ***IEEE Transactions on Vehicular Technology***, vol. 66, no. 9, pp. 8553-8557, Sept. 2017.
- J-9 . X. Lyu, H. Tian. “Adaptive Receding Horizon Offloading Strategy Under Dynamic Environment” in ***IEEE Communications Letters***, vol. 20, no. 5, pp. 878-881, May 2016.

# Contents

Certificate	ii
Abstract	iii
Acknowledgments	iv
List of Publications	v
List of Figures	xi
Abbreviation	xv
<b>1 Introduction</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.1.1 Challenges of Fog Computing in Future Networks . . . . .	2
1.2 Thesis Organization . . . . .	6
<b>2 Literature Review</b>	<b>8</b>
2.1 Fog Computing and its Reference Architecture . . . . .	8
2.2 Resource Management Models in Fog Computing . . . . .	11
2.3 Summary of Related Work . . . . .	12
<b>3 Multiuser Joint Task Offloading and Resource Optimization</b>	<b>16</b>
3.1 System Model . . . . .	17
3.2 Problem Formulation . . . . .	20
3.3 Decomposition-Based Problem Reduction . . . . .	22

3.3.1	Joint Optimization of Communication and Computation Resources . . . . .	23
3.3.2	Submodular Maximization Problem . . . . .	27
3.4	Heuristic Offloading Decision Algorithm . . . . .	28
3.4.1	Locally Optimal Offloading Decision . . . . .	28
3.4.2	User Classification . . . . .	30
3.4.3	Heuristic Offloading Decision Algorithm . . . . .	31
3.5	Evaluation . . . . .	34
3.5.1	Analysis of System Utility . . . . .	36
3.5.2	Analysis of per-user Metrics . . . . .	38
<b>4</b>	<b>Energy-efficient Admission Control of Delay-Sensitive Tasks</b>	<b>43</b>
4.1	System Model . . . . .	44
4.2	Energy-efficient Offloading and Resource Optimization . . . . .	46
4.2.1	Problem Formulation . . . . .	47
4.2.2	Quantized DP Algorithm . . . . .	50
4.3	Asymptotically Optimal Quantization Interval . . . . .	53
4.4	Discussions and Extensions . . . . .	56
4.4.1	Complexity Analysis . . . . .	57
4.4.2	Fairness and Task Partitioning . . . . .	58
4.4.3	Channel-aware Scheduling . . . . .	59
4.4.4	The Impact of Inter-cell Interference . . . . .	61
4.5	Simulation Results . . . . .	62
<b>5</b>	<b>Optimal Schedule for Internet-of-Things under Partial</b>	



<b>Information</b>	<b>71</b>
5.1 System Model . . . . .	72
5.2 Online Offloading in Single-cell Fog Computing . . . . .	75
5.2.1 Perturbed Lyapunov Optimization . . . . .	76
5.2.2 Optimal Offloading Schedules . . . . .	82
5.3 Offloading Schedules under Partial Feedbacks . . . . .	85
5.4 Simulation Result and Analysis . . . . .	87
<b>6 Distributed Fog Computing for Selfish Devices</b>	<b>93</b>
6.1 System Model . . . . .	94
6.2 Distributed Online Optimization of Fog Computing for Selfish Devices	99
6.2.1 Problem Statement . . . . .	99
6.2.2 Distributed Online Operations via Lyapunov Optimization . .	101
6.2.3 Performance Analysis . . . . .	108
6.3 Distributed Optimization under Out-of-Date Knowledge . . . . .	111
6.4 Simulation Result and Analysis . . . . .	116
<b>7 Collaborative Regions for Large-scale Fog Computing</b>	<b>122</b>
7.1 System Model . . . . .	124
7.2 Distributed Online Control of Fog Computing . . . . .	128
7.2.1 Distributed Online Optimization . . . . .	128
7.2.2 Asymptotic Optimality . . . . .	134
7.3 Collaborative Region for Large-scale Inhomogeneous Networks . . . .	137
7.3.1 Adaptive Collaborative Region . . . . .	137
7.3.2 Distributed Formation of Adaptive Collaborative Region . . .	139
7.4 Simulation Result and Analysis . . . . .	141

7.5 Extensions to Unreliable Networks . . . . .	147
<b>8 Conclusion and Future Work</b>	<b>150</b>

## List of Figures

1.1	The application of fog computing in IoT networks. . . . .	3
2.1	The OpenFog reference architecture for fog computing [1]. . . . .	10
3.1	The single-cell multi-user fog computing scenario. . . . .	17
3.2	The comparison of system utility as $K$ varies from 1 to 40. HODA performs very close to the optimal solution. . . . .	36
3.3	The average number of offloaded users as $K$ increases from 1 to 40. Both HODA and the enumeration algorithm offload fewer users as $K$ increases, improving system utility. . . . .	37
3.4	The comparison of CDF of the ratio of the system utility of a solution compared to the enumeration algorithm when $K = 40$ . HODA does not exhibit high variations. . . . .	38
3.5	The comparison of CDF of user utility when $K = 40$ . All users have non-negative utility with HODA. . . . .	39
3.6	The comparison of CDF of task completion time when $K = 40$ . Offloading too many users leads to an increase in task completion time. . . . .	39
3.7	The comparison of CDF of task completion time when $K = 10$ . HODA improves task completion time. . . . .	40
3.8	The comparison of CDF of energy consumption when $K = 40$ . Selectively offloading tasks improves energy efficiency. . . . .	40

3.9	The average latency and energy consumption of HODA as $\beta_i^E = 0.25-0.75$ with $\beta_i^T = 0.5$ when $K = 10, 40$ . Increasing $\beta_i^E$ reduces the energy consumption at the expense of longer latency. . . .	41
4.1	The single-cell multi-user fog computing system with delay-sensitive tasks. . . . .	44
4.2	The $[\mathcal{O}(\varepsilon), \mathcal{O}(1/\varepsilon)]$ -tradeoff between the optimality loss and time-complexity. . . . .	64
4.3	The comparison of devices' average energy consumption as $T^{\text{req}}$ increases from 1s to 3s, where $N = 20$ . . . . .	65
4.4	The comparison of the number of satisfied devices as $f_0$ increases from 10GHz to 30GHz, where $N = 20$ and $T^{\text{req}} = 1\text{s}$ . . . . .	66
4.5	The comparison of devices' average latency and energy consumption where $T^{\text{req}} = 1\text{s}$ . EROS and B&B increase energy consumptions to satisfy the stringent deadline. . . . .	66
4.6	The comparison of devices' average latency and energy consumption where $T^{\text{req}} = 1.5\text{s}$ . EROS and B&B can substantially save energy. . .	68
4.7	The comparison of the numbers of admitted and satisfied devices where $T^{\text{req}} = 1\text{s}$ . Without task admission, ARAA satisfies no task deadlines. . . . .	68
4.8	The comparison of the numbers of pre-admitted and pre-denied devices where $T^{\text{req}} = 1\text{s}$ . The number of devices to be assessed is consistent with the result in Fig. 4.9. . . . .	69
4.9	The comparison of runtime between EROS and B&B where $T^{\text{req}} = 1\text{s}$ . EROS is superior in terms of efficiency and stability. . . .	70
5.1	An illustrative example of single-cell fog computing for IoT. . . . .	72

5.2	The changes of throughput and the corresponding average number of feedbacks as $N$ increases from 100 to 5000 where $V = 40$ . . . . .	89
5.3	The system throughput and Jain's fairness index of the proposed approach, RR and PF, where $V$ ranges from 0 to 40 and $N = 500$ . . .	90
5.4	The comparison of the simulation results on the maximum backlogs of the data and energy queues where $V$ increases from 0 to 40, where $N = 500$ . . . . .	90
5.5	The time variation of the average data backlog of the proposed approach, RR and PF along the time, where $N = 500$ . . . . .	91
5.6	The comparison of throughput between the proposed approach, RR and PF, versus data arrival and CPU resources, where $N = 500$ and $V = 40$ . . . . .	92
6.1	An illustrative example of a distributed wireless network supporting fog computing. . . . .	94
6.2	An illustration on task processing and offloading, and the return of results in the proposed approach. The virtual queues are designed to capture the tit-for-tat incentives which motivate selfish devices to cooperate, as will be introduced in Section 6.2. . . . .	98
6.3	The comparing of the stabilized system throughput and energy efficiency between the proposed approach and the existing benchmarks, as $A_i(t)$ increases from 0.5 to 4.5 Mbits. . . . .	117
6.4	The stabilized system throughput and energy efficiency of the proposed approach versus connectivity ratio, where $V = 70$ and $A_i(t) = 4.5$ Mbits. . . . .	118
6.5	The stabilized system throughput and energy consumption versus $V$ . . . . .	119
6.6	The average backlog of the task and result queues versus $V$ . . . . .	120
6.7	The change of the average backlog over time. . . . .	120

6.8	The change of the backlog of virtual queues over time. . . . .	121
7.1	An example of three servers to illustrate the offloading decision-making over a link at time $t$ . The lengths of the queues, in which server 1 buffers the unprocessed tasks of itself, server 2 and server 3, are $Q_1^{(1)}(t) = 59\text{Mbits}$ , $Q_1^{(2)}(t) = 21\text{Mbits}$ and $Q_1^{(3)}(t) = 18\text{Mbits}$ . The lengths of the queues, in which server 2 buffers the unprocessed tasks of server 1, itself, and server 3, are $Q_2^{(1)}(t) = 13\text{Mbits}$ , $Q_2^{(2)}(t) = 32\text{Mbits}$ and $Q_2^{(3)}(t) = 19\text{Mbits}$ , respectively. Consider the link between servers 1 and 2. The price of offloading between servers 1 and 2 is $\zeta_{12}(t) = 0.5$ . . . . .	123
7.2	An illustration on the proposed admission, processing and offloading of tasks, and the return of results by having all $N$ servers collaborate.	124
7.3	The change of the system cost over time, where $1/\epsilon=50$ . . . . .	142
7.4	The average collaborative region size, as $N$ increases from 10 to 500. .	143
7.5	The system income and throughput versus collaborative region size, where $1/\epsilon = 50$ . . . . .	144
7.6	The ratio of income between Fixed offloading and Proposed and the average region size versus maximum connected servers per node, where $1/\epsilon = 50$ . . . . .	144
7.7	The ratio of income between Fixed offloading and Proposed and the average region size versus the percentage of servers without task arrivals, where $1/\epsilon = 50$ . . . . .	145
7.8	The stabilized system income versus $1/\epsilon$ . . . . .	146
7.9	The total queue backlogs and average region size versus $1/\epsilon$ . . . . .	147

# Abbreviation

Internet of Things - IoT

Quality of service - QoS

Non-deterministic polynomial hard- NP-hard

Mixed integer programming - MIP

Dynamic programming- DP

Time-division multiple access - TDMA

Orthogonal frequency-division multiple access - OFDMA

Alternating direction method of multipliers - ADMM

Quality of experience - QoE

Base station - BS

Long-term evolution - LTE

Mixed integer non-linear programming - MINLP

KarushKuhnTucker - KKT

Cumulative distribution function - CDF

Integer programming - IP

Linear programming - LP

Independent and non-identically distributed - i.n.d.

Independent and identical distributed - i.i.d.

Round robin - RR

Proportional fair - PF

Device-to-Device - D2D

Precision time protocol -PTP

Timing-sync protocol for sensor networks - TPSN

First-in-first-out - FIFO

Left-hand-side - LHS

Right-hand-side - RHS

Neighbor discovery protocol - NDP

Head-of-line - HOL